technical notes Journal of proteome research

Correlation of Multiple Peptide Mass Spectra for Phosphoprotein Identification

Jian Feng,[†] Wesley M. Garrett,[‡] Daniel Q. Naiman,[†] and Bret Cooper*[,§]

Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, Maryland 21218, Animal Biosciences and Biotechnology Laboratory, USDA-ARS, Beltsville, Maryland 20705, and Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, Maryland 20705

Received July 7, 2009

Abstract: When collision induced dissociation is used to fragment phosphorylated peptides during tandem mass spectrometry (MS^2), an ion exhibiting the neutral loss of phosphoric acid can be the major product. The neutral loss ion can then be fragmented during MS^3 for additional resolution of the peptide sequence. Together, MS^2 and MS^3 spectral pairs can offer supporting identification of phosphorylated peptides and proteins. Here, the software program PANORAMICS has been modified to make use of Mascot results for MS^2 and MS^3 spectral data sets. For pairs, the algorithm considers the number of shared *m*/*z* peaks used for peptide assignment and then adjusts the score evaluating that a peptide was correctly matched to these spectra using a mathematical model. The algorithm then calculates peptide probabilities for paired and unpaired spectra, and deduces a probability that a protein was identified given the set of matched peptides. The output provides information useful for determining whether peptides and proteins are phosphorylated. The program can process large result files often generated by multidimensional protein identification technology (MudPIT).

Keywords: phosphorylation • phosphopeptide • neutral loss • MS^3

Introduction

We have been using a hybrid Orbitrap/linear ion trap mass spectrometer for discovering phosphorylated proteins by collision-induced dissociation (CID), as have others.[1−3] In one common configuration, the instrument is set to perform a high-resolution mass scan in the Orbitrap and then fragment particular ions in the linear ion trap and create tandem mass spectra (MS^2). During CID, nonphosphorylated peptides reproducibly fragment along the peptide backbone and produce an MS^2 spectrum that can be interpreted to resolve the amino acid order of the peptide. However, a phosphorylated peptide often primarily releases phosphoric acid (H_3PO_4, but also $HPO_3 + H_2O$) during CID.[4] This can yield a major peak in the spectrum for which the mass-to-charge ratio differential compared to the parent ion is −98, −49, or −32.7 Da, depending on the ion charge (+1, +2, and +3, respectively). Thus, when neutral loss of phosphoric acid competes with b-y dissociation, an MS^2 spectrum can lack a discernible peptide fragmentation ion series, or it can exhibit a fragmentation ion series with a low signal-to-noise ratio.[5] Notwithstanding, phosphoric acid neutral loss can be exploited and the hybrid Orbitrap/linear ion trap mass spectrometer set to perform another round of CID on the neutral loss ion, leading to an MS^3 spectrum that can contain a peptide fragmentation ion series sufficient for peptide sequence identification.[6] The fact that the MS^3 spectrum generation is triggered by the neutral loss ion can be predictive of the presence of a phosphate group in the MS^2 ion. Together, MS^2 and MS^3 spectral information can be used to confirm the existence of a phosphopeptide.

Many laboratories still routinely analyze MS^2 and MS^3 spectra by hand to cross-validate phosphopeptide information.[4,7] However, software can facilitate discovery, and there are several platforms designed to systematically apply an idea inherent to manual inspection—a peptide sequence resolved by the MS^3 spectrum brings added confidence to understanding a peptide ion observed in the MS^2 event. For example, Olsen and Mann designed software for peptide sequence identification that uses an improved score for a Mascot-generated MS^2 spectrum/peptide sequence match probability and a customized probability from a corroborating MS^3 spectrum.[6] Software by Ulintz et al. uses parallel knowledge of the distributions of paired MS^2 and MS^3 spectra to adjust the PeptideProphet-generated probability that the peptide identification is correct.[8] Meanwhile, Jiang et al. combine Sequest scores for peptide sequences matched to MS^2 and MS^3 spectra and then use a target-decoy approach to empirically set a threshold for accepting peptide sequences identified by the summed score.[9] Although different in their approaches, these programs share the common goal of using additional MS^3 information for the identification of phosphorylated peptides.

In this paper, we introduce our approach for improving phosphorylated protein identification by combining paired MS^2 and MS^3 spectral scores. We differentiate our software chiefly by the principle that our primary goal is to identify proteins and their sets of phosphorylated and nonphosphorylated peptides rather than just identify phosphorylated peptides. This concept is best demonstrated through PANORAMICS, our probability-based program that determines the likelihood that

* To whom correspondence should be addressed at 10300 Baltimore Ave., Bldg. 006, Rm. 213, Beltsville, MD 20705. Phone, 301-504-9892; e-mail, bret.cooper@ars.usda.gov.
† The Johns Hopkins University.
‡ Animal Biosciences and Biotechnology Laboratory, USDA-ARS.
§ Soybean Genomics and Improvement Laboratory, USDA-ARS.

10.1021/pr900596u CCC: $40.75
Published on Web 08/31/2009

peptides are correctly assigned to proteins.[10] PANORAMICS processes Mascot results files and has a number of favorable features. Protein identification false-discovery rates that can be deduced are more accurate than other methods.[10] PANORAMICS also reduces redundancy in the data set by coherently distinguishing distinct and shared peptides,[11] enables label-free protein quantification through the summed spectral count method,[12] and operates exceptionally quickly due to a novel data structure.[13]

Through a new programmatic implementation, PANORAMICS[2] analyzes MS^2 and MS^3 spectrum−peptide pairs in a case-by-case manner instead of evaluating all spectrum−peptide matches across all data sets independent of MS^2 or MS^3 acquisition.[14] A new score based on the original Mascot Ions score and the number of matched peaks unique to a corroborating spectrum is created. The new scores for pairs and the original scores for unpaired spectra are used to calculate the confidence measures of peptide identification. Peptide probabilities are then used to compute a probability for protein identification, as before. The output shows the sets of peptides used to identify a protein, the spectral pairings and the varied positioning assignments of phosphate moieties attributed by Mascot.

Experimental Section

Protein Standards. Tryptic digests of bovine apotransferrin (gi|2501351), bovine glutamate dehydrogenase (gi|118533), bovine catalase (gi|84028182) and bovine lactoperoxidase (gi|129823) were prepared according to manufacturer's instructions (Michrom Bioresources, Auburn, CA). A total of 0.5 pmol of each was analyzed. In addition, 2.5 μg of a preparation of recombinant human osteopontin (O4264, Sigma-Aldrich, St. Louis, MO; gi|91206462), 2.5 μg of chicken albumin (A7641, Sigma-Aldrich; gi|129293), 2.5 μg of porcine troponin (T2275, Sigma-Aldrich; gi|73853890), 1 μg of β-casein (gi|30794310) monophosphopeptide (FQpSEEQQQTEDELQDK, Sigma-Aldrich), and 1 μg of β-casein tetraphosphopeptide (RELEELNVPGEIVEpSLpSpSpSEESITR, Sigma-Aldrich) were also analyzed.

Mass Spectrometry. Peptides were separated on homemade 75 μm i.d. fused-silica columns with a 5 μm tip and packed first with reverse phase C18 resin (Aqua, 5 μm, Phenomenex, Torrance, CA) followed by strong cation exchange resin (Luna, 5 μm, Phenomenex).[15] A 12-step elution procedure consisting of stepwise increasing concentrations of salt solution followed by increasing gradients of organic mobile phase was used.[15] Solvent flow was 200 nL/min and was controlled with an Accela HPLC pump (Thermo Fisher Scientific, Waltham, MA) and a T-split junction where 2100 V electricity was applied.[15] The eluent was electrosprayed directly into the orifice of an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) controlled by Xcalibur 2.0.7 software (Thermo Fisher Scientific). A parent-ion scan was performed in the Orbitrap over the range of 400−1600 m/z at 60 000 resolution, with 500 000 automatic gain control (AGC), 500 ms ion injection time, and 1 μscan. Lock-mass was enabled.[16] Data-dependent MS^2 and MS^3 were performed in the linear ion trap with 10 000 AGC and 100 ms ion injection times with 1 μscan. MS^2 was performed on the five most intense MS ions, and MS^3 was triggered if one of the top three MS^2 peaks corresponded with neutral loss of 98.0, 49.0, 32.7 m/z.[17] Minimum signals were 1000 and 100, respectively. An isolation width of 2 m/z and normalized collision energy of 35% were used for MS^2 and MS^3. Dynamic exclusion was used with repeat count of 2, 30 s repeat duration, a list of 100, list duration of 2 min and exclusion mass width of ±0.7 Da. Three separate protein and peptide preparations were analyzed.

Mascot Searching. MS^2 and MS^3 spectrum data files were separately extracted from the raw data with Bioworks 3.3.1 (Thermo Fisher Scientific) using the parameters 600−4500 mass range, 1 group scan, 1 minimum group count, and 5 minimum ion counts. Sets of combined MS^2 spectra and sets of combined MS^3 spectra were searched with Mascot 2.2.[18] For sets of MS^2 spectra, search parameters were for tryptic digests, 1 possible missed cleavage, fixed amino acid modification [+57, C], variable amino acid modifications [+80, S] and [+80, T], monoisotopic mass values, ±10 ppm parent ion mass tolerance, ±0.5 Da fragment ion mass tolerance, and #^{13}C = 1 enabled. For MS^3 spectra, search parameters were for tryptic digests, 1 possible missed cleavage, fixed amino acid modification [+57, C], variable amino acid modifications [+80, S], [+80, T], [−18, S], and [−18, T], monoisotopic mass values, ±1.2 Da parent ion mass tolerance, and ±0.5 Da fragment ion mass tolerance. The searched database consisted of the protein standard records, some of their variants, common contaminants plus 1000 reversed *Saccharomyces cerevisiae* protein sequences (1198 records total).

Ions Score Hypothesis. Mascot result files contain an Ions score for each peptide−spectrum match showing the likelihood of the match to be correct. This score is based on the comparison of m/z peaks between the observed spectrum and the predicted spectrum. Although the details of the Mascot algorithm have not been fully disclosed, we do know the Ions score will be 0 if no m/z peak is matched, or whenever the number of matches is worse than we would expect by chance, and that the score rises when more peaks are matched. We call the matching m/z values "matched peaks". The peaks mentioned in this article correspond to not only the b and y series ions, but also the a, z and other ion peaks, depending on what is measured by the particular instrument and considered by Mascot.

We hypothesize that the Ions score is proportional to the number of matched peaks. To test this, we subtracted matched peaks from a spectrum to see if the Ions score would drop at a proportional rate (Figure 1A). A particular spectrum of a phosphopeptide with a very high Ions score of 115 was chosen to allow us to measure a broader range of scores as peaks were subtracted. Mascot displayed in its html output 30 matched peaks that contributed to the original score (only b and y ions are shown in Figure 1A−C, however). Two pairs of ions from different series appeared to be matched to the same peak, meaning that it was only possible to delete 28 of the matched peaks. Thus, we then deleted the corresponding 28 mass and intensity values (and their ^{13}C isotopes) from the peak list, one by one from highest intensity to lowest and searched each consecutive depleted peak list with Mascot. Upon deletion of the 14th originally matched peak, Mascot stopped interpreting the spectrum using the phosphoric acid neutral loss ion series and started interpreting the spectrum using the ion series without the neutral loss for the same peptide sequence (Figure 1B). At this point, we then deleted the mass and intensity values for the non-neutral loss ion series one by one in addition to the original matched peaks, which prompted Mascot to resume interpreting the spectrum using the prior phosphoric acid neutral loss series. Despite the decreasing scores and the depleted peaks, all peak lists were matched to the same peptide sequence. The reasons why Mascot still correctly matched the

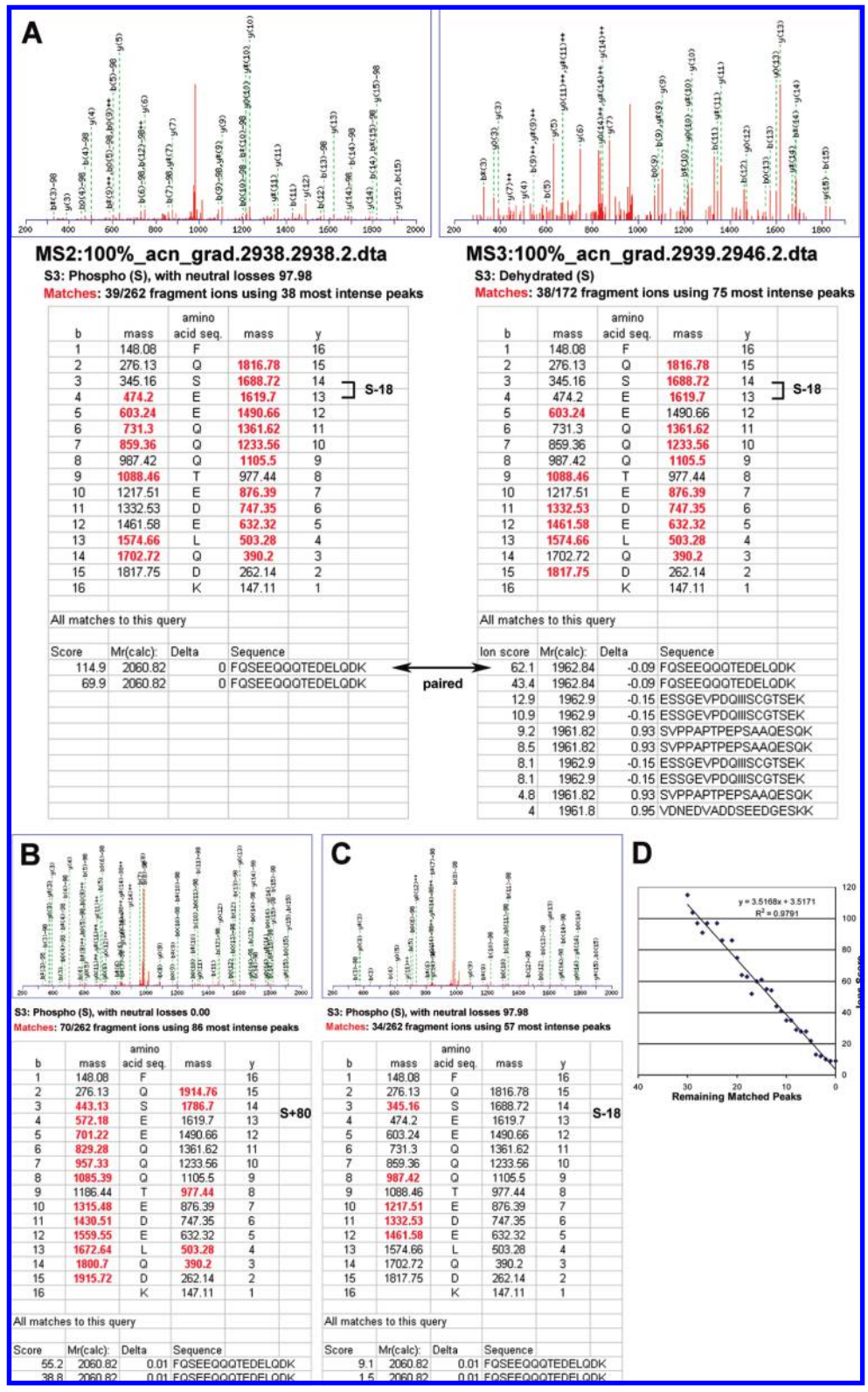

Figure 1. MS^2 and MS^3 spectral pairing, matched b and y ions, and evidence of Mascot Ions score linearity. (A) Consecutive MS^2 and MS^3 spectra for the same peptide are paired and exhibit concordance with respect to the matched peptide sequence, matched peaks and amino acid modification. The *m/z* for the parent ion for the MS^2 spectrum was 1031.42029 and the *m/z* for the parent ion for the MS^3 spectrum was 982.366. Mass shifts and Mascot match pairing noted. The ranking peptide matches are also shown. For both spectra, the second ranking match is to the same peptide sequence as the first but the modification assignment is made to T rather than S. (B) MS^2 spectrum in panel A with top 14 most intense matched peaks deleted. As a result, Mascot no longer interpreted the neutral loss of −98 Da, but instead detected evidence for +80 Da S. (C) Spectrum deleted of all original matched peaks of ion series for neutral loss of −98 Da as well as the ion series for neutral loss of 0.00. (D) Plot of Mascot Ions score vs number of remaining matched peaks originally matched in panel A.

Table 1. Definitions of Terms for the Model for Coupling MS^2 and MS^3 Spectra

L = Lowest m/z ratio detectable for mass spectrometer
H = Highest m/z ratio detectable for mass spectrometer
E = m/z error
A = Spectrum A of peptide R
B = Spectrum B of peptide R
M_1 = no. of peaks in A
M_2 = no. of peaks in B
S_1 = Mascot Ions score for A
S_2 = Mascot Ions score for B
N_1 = no. of peaks that contribute to S_1 (matched peaks for A)
N_2 = no. of peaks that contribute to S_2 (matched peaks for B)
P = no. of matched peaks in B that already are in A and contribute to S_1
$N_2 - P$ = no. of unique matched peaks to B
Q = number of "good" bins that correspond to all the expected peaks in a spectrum matched to a given peptide sequence

peptide sequence to the final depleted spectrum (Figure 1C) are (1) there is only one candidate peptide within the narrow mass range given the specificity of the small database and (2) as the original peaks were deleted, Mascot found other peaks, once likely to be deemed minor, to then be sufficient for matching. Obviously, even with the low Ions score of 9, few major ions from the b and y series were matched in the spectrum in Figure 1C. The graph in Figure 1D shows that the Mascot Ions score decreases linearly in relation to the remaining number of originally matched peaks. These data support our hypothesis.

Model for Coupled MS^2 and MS^3 Spectra. Suppose MS^2 spectrum A and MS^3 spectrum B are both derived from peptide R with Ions Score S_1 and S_2, and with numbers of peaks M_1 and M_2, respectively (see Table 1 for definitions). The method we use to correlate these two spectra can be divided into seven steps:

Step 1: Bin the Peaks in Spectra A and B. Suppose the mass range of the spectrometer is $[L, H]$, and the error is E. We can think of the mass range to be $(H - L)/E$ bins. By definition, the "number of peaks" refers to the number of occupied bins.

Step 2: Designate the Spectrum Whose Ions Score Is Higher As the Base Spectrum. Suppose $S_1 \geq S_2$, so A is chosen as the base.

Step 3: Find How Many Peaks in Spectrum B Are Peaks That Have Contributed to the Ions Score S_2. These are defined as matched peaks. Suppose B has N_2 matched peaks, respectively.

Step 4: Find How Many Matched Peaks in Spectrum B Already Appeared in Spectrum A and Already Contributed to S_1. Suppose this number is P. Then there are $N_2 - P$ peaks in spectrum B that are matched in B but did not appear in spectrum A. These are unique matched peaks to B.

Step 5: Estimate the Portion of the Falsely Matched Peaks Among all the M_2 Peaks. A spectrum may contain hundreds of peaks, a majority of which could be noise. Some of these noise peaks could be used by the database search engine and contribute to the Ions score if they have the same mass-to-charge ratio as some of the fragmented ions within an error range.[18] The exact number of noise peaks that are falsely matched cannot be calculated accurately because of the random nature of appearance of the noise peaks. However, it can be estimated approximately with a basic probabilistic hypothesis. Since there are $(H - L)/E$ small bins in the mass range of the MS instrument, we can calculate how many of them are "good" bins, meaning that an observed peak will be tagged as a matched peak if it falls into these bins. The number of "good" bins, denoted as Q, can be calculated for any given peptide sequence. These "good" bins correspond to all the expected peaks in a spectrum matched to a given peptide sequence, while N_2 is equivalent to the number of bins that correspond to all the observed peaks that are matched to the peptide sequence. Q is usually much greater than N_2 because many expected peaks might not be observed in a real spectrum. Therefore, not every "good" bin is filled out for a particular spectrum. When we observe a total number of M_2 peaks in a spectrum, we need to estimate how many of the peaks among the M_2 peaks would fall into these Q bins only by chance if these M_2 peaks were randomly thrown into the $(H - L)/E$ small bins. These chance peaks could thus mistakenly contribute to the Ions score. A similar problem in probability theory is that we randomly choose M_2 balls in a bag with Q black balls and $[(H - L)/E - Q]$ white balls, and want to know how many white balls will be chosen. The number of white balls that will be chosen is a random variable and can take any value from 0 to M_2 (if both Q and $[(H - L)/E - Q]$ are greater than M_2).The probability distribution of this random variable is called the hyper-geometric probability distribution, and has expectation as

$$M_2 \frac{QE}{H - L}$$

We use this value as the estimate of the number of peaks which are falsely identified as matched peaks. Therefore, the portion of the falsely matched peaks among all the M_2 peaks is

$$\frac{QE}{H - L}$$

Step 6: Estimate How Many of the $(N_2 - P)$ Matched Peaks in B (As Defined in Step 4) That Did Not Appear in A Are Noise. If we assume that the same portion in $(N_2 - P)$ peaks is falsely matched as that in the M_2 peaks, then the estimate is

$$(N_2 - P) \frac{QE}{H - L}$$

Step 7: Update the Ions Score of the Match between the Peptide and the Spectrum A. We separate the matched peaks in spectrum B into two categories: those that also appeared in spectrum A and those did not appear in spectrum A. We then split the score of the peptide matched to spectrum B into two corresponding parts, one that corresponds to the peaks in the first category (i.e., that appeared in A), and one that corresponds to the second category (i.e., that did not appear in A). The score is split proportionally to the number of peaks in the two parts. We then add the second part of the score to S_1, the score of the match from spectrum A to the peptide. Thus, the updated score is

$$S = S_1 + \left((N_2 - P) - (N_2 - P)\frac{QE}{H - L}\right)\frac{S_2}{N_2}$$

The confidence level of peptide identification from spectra A and B can then be derived using the updated Ions score max(S,

S_1) instead of two Ions scores S_1 and S_2. (See Supplemental Data 1 as to why this method of updating the Ions score max(S, S_1) is not the same as merging the MS^2 and MS^3 peak lists and rescoring the combined spectrum with Mascot.)

It should be noted that the adjustment of Ions score from S_1 to max(S, S_1) is conservative because in Steps 5 and 6, when we estimate the number of incorrectly matched peaks, we assume all the observed peaks are noise, which is obviously not true. As a result, the number of incorrectly matched peaks could be overestimated, which in turn drives the value of S down. Furthermore, the "bin" method is only used to estimate the value of the portion at the last part in Step 5 and Step 6. The actual values for ions combined with the instrument's error range are used both for MS^2 and MS^3 spectra to decide if a particular peak is matched.

It is worth elaborating that, when deriving a confidence measure for peptide identifications, an algorithm has to assume a suitable mathematical model to combine the information from the different spectra that are matched to the same peptide sequence, such as the independence probability model.[8,10,19] However, the independence model is problematic if the spectra share a large number of peaks, even though the precursor ions have different charge states, variable modifications, or neutral losses. In such cases, the model assuming independence between the correctness of these matches could greatly overestimate the confidence measure of the peptide to be correctly identified. On the basis of this reasoning, the maximum probability model might seem to be a better choice, because it only chooses the best match among multiple spectra to a single peptide. However, experiments have shown that the same peptide under different circumstances can have very different fragmentation patterns, and therefore, could give rise to very different spectra. Since in this case large amounts of the information contained in the different spectra that identifies the peptide are not congruent, but are independent for the identification of the same peptide, the maximum probability model could underestimate the confidence measure, and the independence model could be more preferable. It is therefore clear that no single probability model can handle these different and complex situations in a consistent way if it does not delve into more details of the relevant spectra but only makes use of the features of precursor ions such as charge states, variable modifications, and neutral losses. We believe a practical start to combine the information in MS^2 and MS^3 spectra matched to a single peptide sequence is to find the number of matched peaks shared by the spectra, and determine how many of these peaks contribute to the Ions scores of each peptide−spectrum match. Our algorithm herein should be considered as one viable but nonexclusive approach.

PANORAMICS² Application. This algorithm was integrated with the original PANORAMICS platform.[10] Only one parametric change was made: for a spectrum to be considered in the calculations, its Ions score − Identity score difference cannot be less than −5. In the first version, the value was −10. The main impact of this adjustment is a reduction in the number of poorer scoring spectra that get evaluated, hence, a reduction in the complexity of the output. No modifications were made to variable parameters previously set by linear regression for calculating probability. Finally, the modified version allows input for two Mascot results files (.dat), one computed for the MS^2 spectra and one for the corresponding MS^3 spectra generated from a data-dependent MS^3 experiment. Although we applied this method to identify peptides with phosphoric acid neutral losses, we expect this method to work equally well on other data-dependent neutral loss generated MS^3 spectra, such as those of oxidized methionine or phosphorylated tyrosine peptides.

Results and Discussion

PANORAMICS² Output Description. We used a hybrid Orbitrap/linear ion trap mass spectrometer to analyze known and unknown phosphorylated and nonphosphorylated peptides. The MS scan was performed in the Orbitrap because there is a clear benefit to peptide identification when using high-mass-accuracy data.[20] We are unaware how the following results might have been impacted had other instruments, methods, or search software been used. These issues are subject to future research.

The hybrid instrument was configured to generate MS^2 and data-dependent MS^3 spectra. We extracted MS^2 and MS^3 spectra separately and searched their peak lists separately to generate two different Mascot results files which were then processed together by PANORAMICS² where information regarding peptide sequences matched to MS^2 and MS^3 spectra were linked by scan numbers. It is possible to concatenate MS^2 and MS^3 peak lists and perform a united Mascot search, but we did not because this requires special modification of the peak lists to instruct Mascot to execute different search parameters depending whether the spectrum is MS^2 or MS^3. In addition, we did not perform a united search because Mascot does not specially consider the relationship between MS^2 and MS^3 pairs or provide a useful way to make this connection if the MS^2 and MS^3 spectra are searched concurrently.

Peptide sequences associated with lone MS^2 and unpaired MS^3 spectra (orphans) were processed alongside the pairs in PANORAMICS² and their derived probabilities were used in conjunction to deduce a protein probability. This deviates from some other approaches where orphans were excluded from analysis[9] and reflects our primary emphasis to derive a program suitable for the identification of proteins and all of their associated peptides regardless of their modification state, as opposed to just identifying phosphorylated peptides.

Example PANORAMICS² data output for purified monophoshorylated and tetraphosphorylated β-casein peptides are shown (Figure 2). Protein group probability and associated protein records are shown first, followed by several rows and columns. A primary row reveals a matched peptide sequence from the protein or group of proteins, the number of spectra assigned to this sequence, whether the sequence is shared or distinct among protein groups, the number of missed cleavages and the probability that a peptide sequence was correctly matched (with respect to the highest-scoring candidate spectra for each particular charge state observed). The secondary rows reveal the modification and neutral loss strings the numbers of which correspond with the variable modification input and the positions of which correspond to the modified amino acid. These strings originate in the Mascot results files and are only apparent to the casual Mascot user once they are decoded as peptide modification features through links in Mascot's html output. Because the links in the Mascot html output are limited to displaying information for one spectrum match at a time, PANORAMICS² extracts this code so that all of this relevant information is visible at once, which provides greater insight into spectrum interpretation. In our data sets, 0 in the modification string = no modification, 1 = phosphorylated S, 2 = phosphorylated T, 3 = dehydrated S, and 4 = dehydrated

1. Protein group probability: 1.0000.	**The equivalent proteins include**		**gi|30794310 beta casein [Bos taurus]**			
Peptides of the group/ 1st spectrum (mod), (neutral loss): paired spectrum (mod), (neutral loss).	spectral count/ Ions score (dash adjusted score)	distinct or shared/ Identity score	missed cleavages/ observed peptide mass(es)	predicted peptide mass(es)	charge state(s)	peptide probaility/spectrum number(s)
FQSEEQQQTEDELQDK	105	distinct	0			1
(0010000000000000), (0010000000000000): (0030000000000000),.	114.94 - 130.43	12.04	2060.82602: 1962.752	2060.821: 1962.844	2	MS2:100%_acn_grad.2938.2938.2.dta MS3:100%_acn_grad.2939.2946.2.dta
(0010000000000000), (0020000000000000): (0030000000000000),.	100.49 - 108.00	13.98	2060.83359: 1962.862	2060.821: 1962.844	2	MS2:100%_acn_grad.2674.2674.2.dta MS3:100%_acn_grad.2675.2675.2.dta
(0000000040000000),: (0000000020000000), (0000000010000000).	25.14 - 38.92	30.01	1962.684: 2060.82871	1962.844: 2060.821	2	MS3:100%_acn_grad.2135.2135.2.dta MS2:100%_acn_grad.2134.2134.2.dta
(0000000000000000),	116.79	13.62	1980.85812	1980.855	2	MS2:100%_acn_grad.3158.3158.2.dta
IEKFQSEEQQQTEDELQDK	403	distinct	1			0.9997
(0000010000000000000), (0000020000000000000)	40.62	14.15	2432.03024	2431.043	3	MS2:10_%_Salt.4980.4980.3.dta
(0000000000020000000), (0000000000020000000)	17.27	13.62	2432.02658	2431.043	3	MS2:50_%_Salt.5144.5144.3.dta
(0000010000000000000), (0000010000000000000)	36.51	14.15	2432.03006	2431.043	3	MS2:20_%_Salt.5059.5059.3.dta
(0000000000020000000), (0000000000010000000)	47.55	14.77	2432.03476	2431.043	2	MS2:20_%_Salt.7210.7210.2.dta
(0000010000000000000), (0000010000000000000): (0000030000000000000),.	60.90 - 63.72	14.77	2432.03476: 2333.880	2431.043: 2333.066	2	MS2:20_%_Salt.7210.7210.2.dta MS3:20_%_Salt.7211.7211.2.dta
(0000000000040000000),: (0000000000020000000), (0000000000010000000).	29.98 - 32.73	30.74	2333.764: 2432.03647	2333.066: 2431.043	2	MS3:20_%_Salt.4862.4869.2.dta MS2:20_%_Salt.4861.4861.2.dta
(0000030000000000000),: (0000010000000000000), (0000010000000000000).	28.10 - 33.04	30.74	2333.764: 2432.03647	2333.066: 2431.043	2	MS3:20_%_Salt.4862.4869.2.dta MS2:20_%_Salt.4861.4861.2.dta
ELEELNVPGEIVESLSSSEESITR	91	shared	0			0.9997
(0000000000000030110000020), (0000000000000000220000010)	41.69	32	2867.227	2867.180	3	MS3:100%_acn_grad.6159.6159.3.dta
(0000000000000010100001040), (0000000000000020200002000)	40.11	32	2867.227	2867.180	3	MS3:100%_acn_grad.6159.6159.3.dta
(0000000000000010100003020), (0000000000000020200000020)	40.11	32	2867.227	2867.180	3	MS3:100%_acn_grad.6159.6159.3.dta
2. Protein group probability: 1.0000.	**The equivalent proteins include**		**gi|129293 ovalbumin [Gallus gallus]**			
FDKLPGFGDSIEAQCGTSVNVHSSLR	73	distinct	1			0.9995
(0000000001000000000000000), (0000000002000000000000000)	20.54	16.02	2900.32229	2900.316	4	MS2:100_%_Salt.7907.7907.4.dta
(0000000001000000000000000), (0000000002000000000000000): (0000000003000000000000000),.	42.03-46.68	15.44	2900.30976: 2803.061	2900.316: 2802.340	3	MS2:90_%_Salt_two.7764.7764.3.dta MS3:90_%_Salt_two.7765.7765.3.dta
EVVGSAEAGVDAASVSEEFRADHPFLFCIK	13	distinct	1			0.9852
(0000100000000000000000000000000), (0000200000000000000000000000000)	31.54	13.22	3317.53073	3316.511	3	MS2:90_%_Salt_one.10783.10783.3.dta
3. Protein group probability: 1.0000.	**The equivalent proteins include**		**gi|912064 osteopontin a [Homo sapiens]**			
FRRPDIQYPDATDEDITSHMESEELNGAYK	5	distinct	1			0.7993
(0000000000020000000000000000), (0000000000010000000000000000)	31.49	13.8	3606.55677	3606.561	3	MS2:70_%_Salt.8568.8568.3.dta
(0000000000000000020000000000000), (0000000000000000010000000000000): (0000000000000000040000000000000),	22.37 - 28.77	13.8	3606.55677: 3509.376	3606.561: 3508.584	3	MS2:70_%_Salt.8568.8568.3.dta MS3:70_%_Salt.8569.8569.3.dta
(0000000000000000010000000000000), (0000000000000000010000000000000): (0000000000000000030000000000000),.	23.47 - 30.98	13.8	3606.55677: 3509.376	3606.561: 3508.584	3	MS2:70_%_Salt.8568.8568.3.dta MS3:70_%_Salt.8569.8569.3.dta
(0000000000000000000000100000000), (0000000000000000000000200000000)	13.36	13.8	3606.55677	3606.561	3	MS2:70_%_Salt.8568.8568.3.dta

Figure 2. Sample output of proteins and identified peptides. Under each protein and its protein probability is the peptide identifying information provided by PANORAMICS[2]. Select peptides and spectra are shown for brevity. The primary rows show a matched peptide sequence from a database, the number of spectra assigned to this sequence, whether the sequence is shared or distinct among protein groups, the number of missed cleavages and the peptide probability. Secondary rows show the modification and neutral loss strings with numbers and positions corresponding to the amino acid modified in the peptide sequence (here, mod. string: 0 = no modification, 1 = phosphorylated S, 2 = phosphorylated T, 3 = dehydrated S and 4 = dehydrated T; Neut. Loss string: 1 = neutral loss of 98 Da was considered at this position, 2 = neutral loss of 98 Da was not considered at this position). The secondary row also shows the Ion score, Identity score, observed peptide mass, predicted peptide mass, ion charge state, and spectrum identification number. Original and adjusted scores are provided for MS^2 and MS^3 pairs.

T at the position indicated (these numbers will change depending on the variety of modifications chosen for any particular search). In the neutral loss string, 1 = spectrum was interpreted such that a neutral loss of 98 Da was considered at the position indicated, whereas 2 = spectrum was interpreted such that a neutral loss of 98 Da was not considered at the position indicated. One interesting feature of Mascot is that neutral loss product ion scanning is automatically engaged once the pS/T variable modification is selected. Therefore, the identification of such an ion dictates the values in the neutral loss string. A second interesting feature is that a modification selection shifts the masses for both the precursor and the modified fragments, whereas a neutral loss shifts only the modified fragments. Thus, if 1 appears in the neutral loss string and a modification of +80 is selected, the net change for the modified fragments is actually −18 Da at the modification position. Note, this is not the same as a selecting for a modification of −18 Da. Likewise, +80 Da is the modified mass differential for the fragments when a 2 appears in the neutral loss string at the same position. Consequently, if a modification string contains the value 1 for S at position x and 3 for S at position y with corresponding neutral loss values of 1 and 0, then S^x is likely dehydrated as a result of the neutral loss of 98 Da at this position, whereas S^y is also likely dehydrated but as a result of some prior event. These "hidden" Mascot interpretations are made apparent by PANORAMICS[2] output and their importance is shown in the following paragraphs.

Continuing with the information in the secondary row, the Ion score, the Identity score, the observed peptide mass, the predicted peptide mass, the ion charge state, and the spectrum identification number are shown. In the event that an MS^2 and MS^3 spectrum are paired, both spectrum identification numbers are shown, and two values for Ions scores are shown: the original and the adjusted score based on information from the pairing. Likewise, if a pairing occurs, two observed and predicted peptide masses are shown, which is useful because, in the event of a true neutral loss, the MS^3 peptide mass should be 98 Da smaller than the MS^2 parent peptide mass. For simplicity, secondary rows show only unique (nonredundant) modification arrangements with respect to the peptide sequence and only the spectrum number (or pairs) with the best score for that arrangement is provided.

Evaluating Lone MS^2 Spectra. PANORAMICS[2] output allows us to observe three basic trends associated with experiments for resolving phosphorylated peptides and proteins, and these are again demonstrated through the β-casein peptide examples. The first trend is the match between an MS^2 spectrum and a peptide sequence but with no supporting MS^3 spectrum (Figure 2, e.g., β-casein, spectrum MS2:20_%_Salt.5059.5059.3.dta). There are several explanations why an MS^2 spectrum does not have a corresponding MS^3 spectrum. Perhaps there were no phosphorylated amino acids or none detected (Figure 2, e.g., β-casein, MS2:100%_acn_grad.3158.3158.2.dta), or there were phosphorylated amino acids but there was no neutral loss to trigger an MS^3 event (Figure 2, e.g., β-casein, MS2:10_%_Salt.4980.4980.3.dta). It is also possible that MS^3 was triggered but that the ions were poorly resolved and screened by filters provided by Xcalibur operating software, Bioworks extraction software or Mascot database-search software. For a lone MS^2 spectrum, the peptide probability that is calculated is solely dependent upon the Mascot Ions score, the Identity score, the database size, the different ion charge states and whether the peptide is distinct or shared. Variables specifically affecting the Ion score are whether or not Mascot interprets the spectrum with or without neutral losses and the total number of potential neutral loss positions with respect to the number predicted by the differential between the molecular weight of the observed parent ion and the theoretical molecular weight of the peptide string. We reiterate that an MS^2 spectrum of a phosphorylated peptide may be interpreted as either having a neutral loss of 98 Da or not, regardless of the generation or presence of a corresponding MS^3 spectrum. This is distinguished by the [1,2] in the neutral loss string output displayed by our software. This has important ramifications that will be discussed later.

In addition to showing that a peptide is phosphorylated, the output shows that Mascot can make numerous interpretations of the positioning of phosphorylation, with each unique interpretation having potential to have a high Ions score (Figure 2, e.g., β-casein, MS2:20_%_Salt.7210.7210.2.dta with a single phosphorylation event predicted at S with Ions score 60.90 or T with Ions score 47.55). Under these circumstances, PANORAMICS[2] considers all of the allowable scores provided for a spectrum or a set of spectra matched to the same peptide sequence and calculates a single peptide probability using the highest calculated probability for each particular charge state.[10] In terms of the model, low scores have low probability and little impact on a high protein probability, whereas high scoring spectra have a greater influence on a protein probability being high. To be clear, the probabilities generated are a product of matching a peptide sequence to a spectrum and do not denote a probability that specially resolves phosphorylation positioning. However, the reporting of the modification position (as viewed by the modification string) with respect to the neutral loss string may be useful when subsequently trying to determine site localization or for determining if localization is ambiguous given the data.

Evaluating Orphan MS^3 Spectra. The second trend that the output allows us to observe is the alternative situation where a peptide is matched to an MS^3 spectrum not paired with an MS^2 spectrum (Figure 2, e.g., β-casein, spectrum MS3:100%_acn_grad.6159.6159.3.dta). Again, there are several explanations for MS^3 orphans, the most likely being that the MS^2 was poorly resolved and did not fulfill the criteria imposed by filters in the various software platforms. In this case, the MS^3 spectrum can impart novel protein sequence identification information not resolved by the MS^2 spectrum. Hence, in PANORAMICS[2], an orphan MS^3 spectrum contributes independently to protein identification in the same manner as a lone MS^2 spectrum, and peptide probabilities derived from orphan MS^3 spectra are contingent upon the same factors as MS^2 spectra. We note that a multiply phosphorylated peptide could also encounter a second neutral loss during the MS^3 fragmentation event and that Mascot can consider this possibility as long as the variable modifications of pS/T are selected for searches with MS^3 spectra (Figure 2, e.g., β-casein, spectrum MS3:100%_acn_grad.6159.6159.3.dta). Thus, just as with MS^2 spectra, MS^3 spectra could also be interpreted to represent a phosphorylated peptide and display a mass shift of −18 Da at S/T if the neutral loss were evaluated or +80 at S/T if the neutral loss were not evaluated. Hence, the presence of [1,2] in the modification string indicates that dehydration at S or T was likely a result of the loss of phosphoric acid during the MS^3 fragmentation event, whereas the [3,4] indicates dehydration as a result of some prior event, possibly the neutral loss of phosphoric acid during MS^2 fragmentation.[21]

Evaluating Paired MS^2/MS^3 Spectra. The third trend is the situation where an MS^2 peptide sequence is supported by the presence of a neutral-loss-generated MS^3 spectrum and its associated peptide sequence, or vice versa (Figure 2, e.g., β-casein, MS2:100%_acn_grad.2938.2938.2.dta and MS3: 100%_acn_grad.2939.2946.2.dta; Figure 1A). An adjusted Ions score is calculated based on the corroborating information from the pairs and this can result in increased peptide identification probabilities. For pairing to occur, their scan numbers must be adjacent. For their scores to be adjusted, their assigned peptide sequences must also be identical. This is important because each spectrum is associated with as many as 10 possible peptide sequence matches (in Figure 1A, the MS^2 spectrum has 2 sequence matches and the MS^3 spectrum has 10). As result of this requirement, the scores for spectra with nonidentical sequence matches are not adjusted but the scores for identical matches are. This allows the situation where a second-best scoring peptide sequence for an MS^3 spectrum can support the top scoring peptide sequence for an MS^2 spectrum (or any other for that matter) as long as the sequences are the same. This is a desirable feature because the top ranking peptide sequence selected by Mascot (or Sequest for that matter) does not always correctly reflect the most accurate interpretation of a spectrum of a phosphopeptide, but a lower-ranking peptide does.[5] To make pairing even more rigorous, the program requires MS^2 peptides to have modifications at the same amino acid positions as MS^3 peptides. The rule guarantees concordance between related ion series in MS^2 and MS^3 spectra. Thus, information from consecutive spectra, where one is produced from the detection of a neutral loss ion, can be used to support each other and increase the confidence of the identification of a modified peptide sequence. Figure 1A is an illustrated example of paired spectra and their common set of matched peaks.

Evaluating Standard Proteins. For the previous examples, we examined spectra from standard β-casein peptides to show the workings of PANORAMICS[2] and its output. To further evaluate the program, we studied a mixture of 7 standard proteins. All but apotransferrin were presumed to possibly be phosphorylated based on information at www.phosphosite.org. Also, the human osteopontin we tested was expressed in mouse culture cells, so it is possible that the phosphorylation state of the protein varied. With PANORAMICS[2], we analyzed Mascot results for 335 507 MS^2 spectra and 17 830 MS^3 spectra separately and together. The MS^2 spectra by themselves were sufficient for identifying all protein standards with a protein probability of 1.0000 (Supplemental Data 2). Identification confidence was mainly due to a combination of high Ions scores and the identification of multiple peptides. The MS^2 spectra also revealed phosphorylated peptides in ovalbumin, osteopontin, and catalase. Altogether, some of the MS^2 data appeared to be adequate to identify peptides and show that the peptides were phosphorylated.

When the Mascot results for MS^2 and MS^3 spectra were analyzed together by PANORAMICS[2], 1 additional peptide contributed to the total peptide count each for glutamate dehydrogenase, apotransferrin, lactoperoxidase and ovalbumin protein identification (Supplemental Data 2). The additional peptides were found with MS^3 orphan spectra. Hence, MS^3 spectra can add to the total protein coverage to improve protein identification. In addition to the orphans, there were MS^3 spectra that correlated with MS^2 spectra. These MS^3 spectra corroborated the MS^2 spectra and increased phosphopeptide identification confidence.

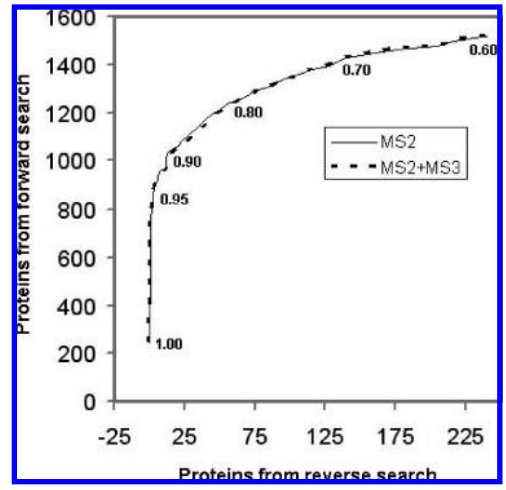

Figure 3. Comparison between the number of proteins identified with PANORAMICS[2] from searches against forward and reverse sequence versions of the *A. thaliana* genome for *A. thaliana* MS^2 spectra and MS^2 + MS^3 spectra. Every point in the curves corresponds to the same protein probability threshold. Some of the probability thresholds are noted next to the curves.

Evaluating Complex Mixtures of Unknown Composition. We present two examples of implementing our software to analyze data from nonstandard samples. A total of 450 μg of soluble protein isolated from a crude membrane fraction from leaves of *Arabidopsis thaliana*[11] was digested with trypsin and the peptides were subjected to the same MudPIT procedure as the standards. The experiment yielded 164 172 MS^2 and 2130 MS^3 spectra, which were searched against the *A. thaliana* genome V 8.0 set of protein sequences using the same parameters as before. PANORAMICS[2] executed its operations on the two Mascot results files in 60 s. A total of 859 nonredundant proteins were found to exceed a 95% protein probability threshold (Figure 3). Fifteen of those were phosphorylated proteins, as evidenced by the detection of phosphorylated S or T on 16 peptides. Two of those identifications were supported by additional MS^3 spectra. A version of the *A. thalania* set of proteins with reversed sequences was also searched and the data was run through PANORAMICS[2]. Only 2 false proteins comprising 1 false peptide match each exceeded the 95% probability threshold. This agrees with our probability estimates which say that, at a 95% probability threshold for data searched against a forward database, less than 5% are likely to be falsely identified. According to the receiver-operator characteristic (ROC) plots for MS^2/MS^3 data analyzed together and MS^2 data analyzed by itself, the accuracy of the PANORAMICS[2] probability model across a range of probabilities remained consistent in light of the increased Ions scores attributed to corroborating MS^3 spectra (Figure 3).

Similarly, we evaluated an IMAC-enriched soybean nucleus lysate. Nuclei were isolated by passing leaf homogenate through different sized mesh screens, extracted in high salt and the chromatin removed by centrifugation.[11] A total of 500 μg of protein was digested and peptides were passed over gallium spin columns.[22] Peptides were eluted with 4 washes of phosphoric acid,[22] pooled and analyzed. In total, 12 144 MS^2 and 172 MS^3 spectra were searched against a *Glycine max* protein sequence database (Glyma1.pep.fa from www.phytozome.net;

75 781 records). In total, 817 proteins exceeding the 95% confidence level were found, while the reverse database search yielded no false proteins exceeding the 95% probability threshold (there were 3 exceeding 90%). 36 phosphoproteins were identified by 36 phosphorylated peptides, 21 of which were supported by paired MS^2 and MS^3 spectra (Supplemental Data 3). By contrast, only 30 phosphorylated proteins/peptides were identified by MS^2 spectra alone. Thus, there is a *bona fide* benefit to coupled MS^2/MS^3 analysis as this can lead to the additional identification of phosphopeptides. For example, a phosphorylated peptide for histone H1/H5 (Glyma02g41220.1) was identified solely as a result of the pairing. The Ions score for the MS^2 spectrum alone was only sufficient for a 72.21% peptide identification probability, while the Ions score for the MS^3 spectrum alone (58.75) was lower than its Identity score (59.36), which may have precluded its use for protein identification had standard Mascot scoring cutoffs been applied. However, after consideration of peptide-determining matched peaks between the dependent spectra, PANORAMICS^2 adjusted the Ions score to 65.90, thereby raising the peptide identification probability to 98.75%. This peptide in addition to another nonphosphorylated peptide contributed to the final identification of the protein at a 99.83% confidence level. This demonstrates an improvement over applying Mascot scoring to evaluate MS^2 and MS^3 spectra independently.

On Evaluating Sites of Phosphorylation. While our program does make use of sets of matched peaks between MS^2 and MS^3 spectra similarly to other programs that specifically evaluate sites of phosphorylation,[6,23] PANORAMICS^2 is not currently designed to predict a probability of certainty for phosphorylation site localization. However, this should not be construed as a deficiency. Palumbo and Reid have provided evidence for CID gas-phase transfer of a phosphate group between a native site and a previously unmodified site, followed by neutral loss of phosphoric acid from the non-native site.[21] They also showed that a phosphorylated peptide may undergo concomitant neutral losses of HPO_3 from a phosphorylated amino acid and H_2O from a nonphosphorylated S or T during MS^2 fragmentation.[21] When these rearrangements or competing fragmentation reactions occur, the resultant MS^2 or MS^3 spectra could yield product ions from which an incorrect phosphorylation site assignment could be made. Since the frequency of rearrangement has not yet been modeled, the accuracy of any probability or scoring system that attempts to pinpoint the sites of phosphorylation using CID MS^2 and MS^3 peptide spectra, but does not consider gas-phase chemistries, is doubtful.[23−26]

However, this does not mean that CID is not useful for detecting phosphopeptides. We have already shown that CID MS^3 can help identify phosphorylated peptides not clearly resolved by MS^2. Furthermore, site positioning can be ascertained by CID if there are no other competitive sites for rearrangement in the peptide, or if phosphate-related neutral losses do not occur. Thus, for the former, if there is only one potential site of phosphorylation in the peptide sequence display of PANORAMICS^2, then the site positioning can be reasonably ascertained, especially if there is any linked, corroborating MS^3 information. For the latter, because PANORAMICS^2 displays the neutral loss sequence string, evaluations of phosphorylation sites in peptide sequences with mod [1,2] and neutral loss [2] can be made (i.e., there was phosphorylation but no neutral loss in that position). Sometimes these +80 modification ion series will produce low Ions scores because they are minor (Figures 1B and 2, osteopontin spectrum MS2:70_%_Salt.8568.8568.3.dta, neutral loss string [2]). However, observed in conjunction with other information for that spectrum and information for related spectra for a particular peptide in question, PANORAMICS^2 may enable improvement in site determination.

Conclusion

PANORAMICS^2 allows the identification of phosphorylated proteins based on their peptide sequences having been matched to MS^2 and MS^3 spectra, with special emphasis being placed on experimentally derived MS^2/MS^3 pairs. This correlation increases the probability that a phosphorylated peptide was identified. The peptide probabilities are factored into the final protein identification probability output. To be clear, PANORAMICS^2 only displays various valid interpretations made by Mascot with regard to the position of a phosphate moiety and does not predict which interpretation is more likely than another. We caution the user who blindly uses Mascot position assignments to make such an assessment.

Although our IMAC-enriched sample served the purpose of demonstrating the effects of our software on a large, nonstandard data set, it has not gone unnoticed by us that the sample did not appear to be overly enriched for phosphopeptides. It is possible that (1) phosphopeptides were not abundant in our nuclear samples; (2) our IMAC methodology is not proficient (although it has been optimized[22] and did yield a greater number of MS^2/MS^3 pairs); and/or (3) PANORAMICS^2 processing of Mascot data returns results differently than other software approaches (to be expected). As for the last possibility, while some programs may accept lower-scoring phosphopeptide identifications and then evaluate only at a peptide level, PANORAMICS^2 is protein-centric, and in order for a high protein probability to be calculated, it may require a higher-scoring unmodified peptide alongside a lower-scoring phosphopeptide. Without additional unmodified peptides, which may be depleted in IMAC-enriched samples, some lower-scoring phosphopeptides may never be observed at a protein probability level. So what program is better? Since peptide- versus protein-centric views are a matter of perspective, it is hard to say. In terms of validated output, we have not yet compared PANORAMICS^2 to any other software that also considers MS^2 and MS^3 pairs.[6,8,9,27] Such a performance test of all of these experimental models will require the diligent establishment of empirical standards and a range of controlled experiments.

Data Availability. PANORAMICS^2 executables for Linux and Windows and open source-codes are freely available in the Supplemental Data.

Acknowledgment. While not trying to imply any endorsement by these people, we simply thank Drs. A. I. Nesvizhskii, S. P. Gygi, D. Tabb, J. Cottrell and G. Reid. This work was funded in part by a specific cooperative agreement between Johns Hopkins University and USDA-ARS and a grant to B.C. from USDA-CSREES (no. 2005-35605-15392).

Supporting Information Available: Supplemental Data 1, data refuting the hypothesis that the PANORAMICS^2 model is the same as merging the MS^2 and MS^3 peak lists (pdf file). Supplemental Data 2, summary of identifying information for standard proteins (pdf file). Supplemental Data 3, PANORAMICS^2 output for IMAC sample (xls file). Supplemental

Data 4, PANORAMICS[2] source-code and executables for Linux and Windows. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312* (5771), 212–217.

(2) Paradela, A.; Albar, J. P. Advances in the analysis of protein phosphorylation. *J. Proteome Res.* **2008**, *7* (5), 1809–1818.

(3) Li, X.; Gerber, S. A.; Rudner, A. D.; Beausoleil, S. A.; Haas, W.; Villen, J.; Elias, J. E.; Gygi, S. P. Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. *J. Proteome Res.* **2007**, *6* (3), 1190–1197.

(4) Lehmann, W. D.; Kruger, R.; Salek, M.; Hung, C. W.; Wolschin, F.; Weckwerth, W. Neutral loss-based phosphopeptide recognition: a collection of caveats. *J. Proteome Res.* **2007**, *6* (7), 2866–2873.

(5) Yang, F.; Stenoien, D. L.; Strittmatter, E. F.; Wang, J.; Ding, L.; Lipton, M. S.; Monroe, M. E.; Nicora, C. D.; Gristenko, M. A.; Tang, K.; Fang, R.; Adkins, J. N.; Camp, D. G., II.; Chen, D. J.; Smith, R. D. Phosphoproteome profiling of human skin fibroblast cells in response to low- and high-dose irradiation. *J. Proteome Res.* **2006**, *5* (5), 1252–1260.

(6) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (37), 13417–13422.

(7) Yu, L. R.; Zhu, Z.; Chan, K. C.; Issaq, H. J.; Dimitrov, D. S.; Veenstra, T. D. Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J. Proteome Res.* **2007**, *6* (11), 4150–4162.

(8) Ulintz, P. J.; Bodenmiller, B.; Andrews, P. C.; Aebersold, R.; Nesvizhskii, A. I. Investigating MS2-MS3 matching statistics: A model for coupling consecutive stage mass spectrometry data for increased peptide identification confidence. *Mol. Cell. Proteomics* **2007**, *7* (1), 71–87.

(9) Jiang, X.; Han, G.; Feng, S.; Jiang, X.; Ye, M.; Yao, X.; Zou, H. Automatic validation of phosphopeptide identifications by the MS2/MS3 target-decoy search strategy. *J. Proteome Res.* **2008**, *7* (4), 1640–1649.

(10) Feng, J.; Naimain, D. Q.; Cooper, B. A probability model for assessing proteins assembled from peptide sequences inferred from tandem mass spectrometry data. *Anal. Chem.* **2007**, *79* (10), 3901–3911.

(11) Lee, J.; Garrett, W. M.; Cooper, B. Shotgun proteomic analysis of Arabidopsis thaliana leaves. *J. Sep. Sci.* **2007**, *30*, 2225–2230.

(12) Lee, J.; Feng, J.; Campbell, K. B.; Scheffler, B. E.; Garrett, W. M.; Thibivilliers, S.; Stacey, G.; Naiman, D. Q.; Tucker, M. L.; Pastor-Corrales, M. A.; Cooper, B. Quantitative proteomic analysis of bean plants infected by a virulent and avirulent obligate rust fungus. *Mol Cell Proteomics* **2009**, *8* (1), 19–31.

(13) Feng, J.; Naiman, D. Q.; Cooper, B. Combined dynamic arrays for storing and searching semi-ordered tandem mass spectrometry data. *J. Comput. Biol.* **2008**, *15* (4), 457–468.

(14) Lu, B.; Ruse, C.; Xu, T.; Park, S. K.; Yates, J. 3rd, Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal. Chem.* **2007**, *79* (4), 1301–1310.

(15) Florens, L.; Washburn, M. P. Proteomic analysis by multidimensional protein identification technology. *Methods Mol. Biol.* **2006**, *328*, 159–175.

(16) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010–2021.

(17) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (33), 12130–12135.

(18) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551–3567.

(19) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75* (17), 4646–4658.

(20) Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villen, J.; Gygi, S. P. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **2006**, *5* (7), 1326–1337.

(21) Palumbo, A. M.; Reid, G. E. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS(3). *Anal. Chem.* **2008**, *80* (24), 9735–9747.

(22) Aryal, U. K.; Olson, D. J.; Ross, A. R. Optimization of immobilized gallium (III) ion affinity chromatography for selective binding and recovery of phosphopeptides from protein digests. *J. Biomol. Tech.* **2008**, *19* (5), 296–310.

(23) Wan, Y.; Cripps, D.; Thomas, S.; Campbell, P.; Ambulos, N.; Chen, T.; Yang, A. PhosphoScan: a probability-based method for phosphorylation site prediction using MS2/MS3 pair information. *J. Proteome Res.* **2008**, *7* (7), 2803–2811.

(24) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, *24* (10), 1285–1292.

(25) Tanner, S.; Payne, S. H.; Dasari, S.; Shen, Z.; Wilmarth, P. A.; David, L. L.; Loomis, W. F.; Briggs, S. P.; Bafna, V. Accurate annotation of peptide modifications through unrestrictive database search. *J. Proteome Res.* **2008**, *7* (1), 170–181.

(26) Ruttenberg, B. E.; Pisitkun, T.; Knepper, M. A.; Hoffert, J. D. PhosphoScore: an open-source phosphorylation site assignment tool for MSn data. *J. Proteome Res.* **2008**, *7* (7), 3054–3059.

(27) Du, X.; Yang, F.; Manes, N. P.; Stenoien, D. L.; Monroe, M. E.; Adkins, J. N.; States, D. J.; Purvine, S. O.; Camp, D. G., II.; Smith, R. D. Linear discriminant analysis-based estimation of the false discovery rate for phosphopeptide identifications. *J. Proteome Res.* **2008**, *7* (6), 2195–2203.

PR900596U